Dixon, T. D., Noyes, J. M., Troscianko, T., Nikolov, S. G., Lewis, J. J., Canga, E. F., ... Canagarajah, C. N. (2007). Scanpath assessment of visible and infrared side-by-side and fused video displays. In 10th International Conference on Information Fusion, 2007, Quebec. (pp. 1 - 8). Institute of Electrical and Electronics Engineers (IEEE). 10.1109/ICIF.2007.4408175

Scanpath Assessment of Visible and Infrared Side-by-Side and Fused Video Displays

T.D. Dixon, J. Li, J.M. Noyes, T. Troscianko
Department of Experimental Psychology
University of Bristol
Bristol, UK
Timothy.Dixon@bristol.ac.uk

S.G. Nikolov, J.J. Lewis, E.F. Canga, D.R. Bull,
C.N. Canagarajah
Centre for Communications Research
University of Bristol
Bristol, UK
Stavri.Nikolov@bristol.ac.uk

Abstract* - Advances in fusion of multi-sensor inputs have necessitated the creation of more sophisticated fused image assessment techniques. The current work extends previous studies investigating participant accuracy in tracking individuals in a video sequence. Participants were shown visible and IR videos individually and the two video inputs side-by-side, as well as averaged, discrete wavelet transform, and dual-tree complex wavelet transform fused videos. Two scenarios were shown to participants: one featured a camouflaged man walking down a pathway through foliage and across a clearing; the other featured several individuals moving around the clearing. The side-by-side scanpath data were analysed by studying how often participants looked at the visible and infrared sides, and analysing how accurately participants tracked the given target, and compared with previously analysed data. The results of this study are discussed in the context of wider applications to image assessment, and the potential for modelling human scanpath performance.

Keywords: Image Fusion, Video Fusion, Scanpath Analysis, Video Assessment, Eye-Tracking, Psychophysics, Side-by-Side Displays, Adjacent Displays

1 Introduction

Previous work involving the novel use of an eye-tracking paradigm to analyse participants' scanpaths, compared visible (Viz) and infrared (IR) individual inputs with averaged (AVE), discrete wavelet transform (DWT), and dual-tree complex wavelet transform (DT-CWT) fused displays [1, 2]. The present study extends this work by introducing a side-by-side (SBS) or adjacent display comprising both the Viz and IR inputs shown simultaneously. The current section presents background research in this area, including the fusion and SBS schemes, whilst Section 2 reports the current experimental method. Section 3 considers the results obtained and compares the SBS results with the previously obtained data, whilst Section 4 is a general discussion.

1.1 Side-by-Side Displays

A wide variety of displays use multiple sources of information, such as IR and Viz light radiation, and present this information to a user in parallel. Some of the simplest and most commonly used 2-D multi-source or multi-modal image displays position the displays either SBS or adjacent, with or without a linked cursor, a 'chessboard' or 'checkerboard' display and a transparency/opacity weighted display. Adjacent or SBS displays present several images simultaneously on the screen (as in the present study), sometimes in multiple windows or on multiple monitors. This is a simple method, but one of the most effective for integrated 2-D displays, especially when a linked cursor is added to assist the observer in relating corresponding features in the images [3, 4]. For example, head mounted displays used by pilots have utilised a dual display with forward-looking IR and a separate night vision goggle sensor that can be activated. However, such systems have been shown to lead to a more time-consuming and confusing visual experience, which can be difficult and distracting for the user [5].

In medical imaging there has also been a strong tradition of assessing images such as positron emission tomography (PET) and computed tomography (CT) scans side by side, in order to compare the complementary information gained from the physiological (PET) and anatomical (CT) images. However, it has been shown that combining these two separate displays into one output that contains the most relevant information of the two inputs can lead to more accurate identification of diagnostically important markers than the SBS images [6].

This paper extends this type of research by focusing on the use of SBS video displays of differing modality inputs (Viz and IR). It investigates when participants decide to look at either Viz or IR, and which is more successful for tracking an individual in a multi-sensor video sequence. In addition, the SBS data is compared with previous results for the Viz and IR both individually, and for three fused displays.

1.2 Image and Video Fusion

Image fusion is the process of combining information from multiple images of a scene, e.g. captured by different sensors, into a single composite image that is more suitable for visual perception or computer processing. There are several benefits to multi-sensor image fusion including wider spatial and temporal coverage, extended range of operation, decreased uncertainty, improved reliability and increased robustness of system performance. Whilst much academic research has recently focused on the methods for fusing static images, and for the assessment of such images, little work has been carried out with regard to multi-sensor video fusion [1, 2, 7].

Two methods of static image fusion that have recently been of interest are the Discrete Wavelet Transform (DWT) and the Dual-Tree Complex Wavelet Transform (DT-CWT). These methods involve transforming the input images into the wavelet domain, with the wavelet coefficients processed and combined based on some fusion rule, and the inverse transform being carried out. The shift-variant DWT method [8] is widely used, for image fusion and is the most basic of the wavelet transform fusion methods.

The DT-CWT [9] method is an alternative form of a DWT. This method has greater directional selectivity than the DWT and is shift invariant with reduced over completeness. DT-CWT has been shown to produce better results in terms of image fusion than other wavelet methods [10, 11], as well as other pyramid and averaging methods [12], across a range of qualitative and quantitative assessments. These advantages come however at the cost of greater computational expense. In the current paper, a simple averaging scheme (AVE) was also used, for reference.

All of these fusion methods can be applied quite simply to videos. One process involves taking each frame individually from a registered sequence of IR-visible video, and fusing each colour plane of the visible sequence separately with the IR sequence. This can then provide a basic, colour-fused output.

1.3 Image Assessment Methods

Recent work has begun to look at objective quantitative ways of human image assessment. Initiated by Toet and colleagues [13, 14], major advances have been made in applying some form of task to the assessment process, and moving away from the ever-present subjective quality assessment. Furthermore, in recent findings [10-12], it has been shown that objective task results can differ significantly from subjective ratings. It is thus essential to choose a well-defined and relevant task when assessing fused images or video sequences, and to go beyond simply applying a subjective rating to the fused outputs.

1.3.1 Eye Tracking Paradigm

One alternative method of attaining data related to visual input is to record scanpaths with the use of eye-tracking technology. A range of eye movements can be found to occur under varying circumstances, including saccades, smooth pursuit, slow drift and stabilisation reflex [15]. Fixation location and length can be an indicator of attention, whilst larger saccades can indicate greater cognitive and perceptual load [16]. The kinds of eye movements that are elicited by a particular task can thus reveal information about the underlying cognitive processes in action.

Investigations into eye movements have considered viewing strategies for people studying complex natural and computer-generated scenes. Individuals have been found to be able to grasp the ‘gist’ of a natural scene very quickly, i.e. within 100ms [17]. Studies considering smooth pursuit eye movements, i.e. those steady movements associated with slow and even tracking, have also found significant variation. It has also been found that the application of a secondary task whilst carrying out smooth pursuit tracking of a target can significantly degrade the performance of the pursuit [18]. Given the broad range of findings, it seems appropriate to apply this knowledge to the area of fused image assessment.

Recent work has also begun to examine fused image quality through the use of scanpath analysis. Krebs and colleagues [19] used such a paradigm to support research into target detection in static scenes. Different fusion methods have been shown to lead to increased and longer fixations in a target detection task, with a principal components fusion scheme leading to greatest number of fixations (whether this necessarily a positive feature of the fusion scheme is not fully explored) [20]. Other work [1, 2] has shown that different fusion methods can be shown to aid participants to track better figures through video sequences, with the simple AVE fusion being particularly useful in this respect.

1.4 Our Approach

The current paper focuses on analysing gaze fixation data in four tracking scenarios using multi-sensor surveillance video data: two simple (SIM) tracking scenarios involving a lone person walking through foliage, down a pathway, and across a clearing of trees; and two complex (COM) scenarios, in which several people interact in a clearing of trees. One simple scenario was filmed at high luminance levels (SIM-HL), whilst the other was filmed at low luminance (SIM-LL), as was the case with the complex scenarios (COM-HL; COM-LL).

2 Method

Parts of the experimental method have previously been published elsewhere [1, 2], and more details of the exact methodology can be found in these papers.

2.1 Design

The current study analysed the pseudo-experimental independent variable (IV) of display type, i.e. whether participants were looking at either the Viz or IR SBS display. Participants were also shown the Viz, IR, and the fused AVE, DWT and DT-CWT displays separately in the same viewing session, although these will currently only be used for comparison purposes with the Viz+IR SBS display. Half the participants were shown the sequences in the order Viz, IR, SBS, AVE, DWT, DT-CWT, and half were shown the reverse order to counterbalance any learning effects. In addition, participants were tested across three sessions, which were treated as a second IV, and the videos were split up into separate sections to aid analysis (a third IV). The dependent variables were the overall percentage of time spent looking at either the Viz or IR display, as well as the accuracy at tracking the target in each display. To compare accuracy of the Viz+IR with the other displays, the accuracy scored in Viz and IR separately was summed providing a total accuracy for the SBS display.

2.2 Participants

Ten participants (5 females and 5 males) took part in the current study. Eight were naïve to the concepts and videos utilised. Ages ranged from 21 to 41 years (mean = 27.1, s.d. = 6.76). Participants had normal or corrected-to-normal vision, and no colour vision problems.

2.3 Apparatus and Stimuli

A Tobii™ x50 remote eye tracker [21] was used to collect eye movement data. This is a table-mounted eye tracker that works at 50 Hz with an approximate accuracy of 0.5°. This was run using the ClearView 2.5.1 software package, on a 2.8 GHz Pentium IV dual processor PC, with 3 GB RAM, and twin SCSI hard drives. Stimuli were presented on a 19" flat screen CRT monitor running at 85 Hz, with screen resolution set to 800 by 600 pixels. Participants used a chin-rest positioned 57cm from the monitor screen.

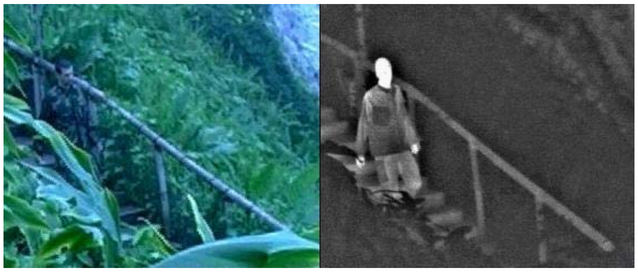

Figure 1: SIM-HL Viz and IR SBS display

The two video sequences shown were part of a data-gathering project carried out at the Eden Project Biome in Cornwall, UK, and detailed in [22]. This project utilised an array of different sensors across two mornings and two evenings filming a variety of scenarios. The four selected for the current paper were subclips from the 'Tropical Forest' collection sequences 2.1_i (SIM-HL), 2.1_iii (SIM-LL), 4.1_i (COM-HL) and 4.1_iii (COM-LL). The SIM sequences both showed a 'soldier' (actor) dressed in camouflaged clothing walking down a pathway amongst foliage, through a clearing of trees and back across the way he came, as shown in Figures 1 and 2.

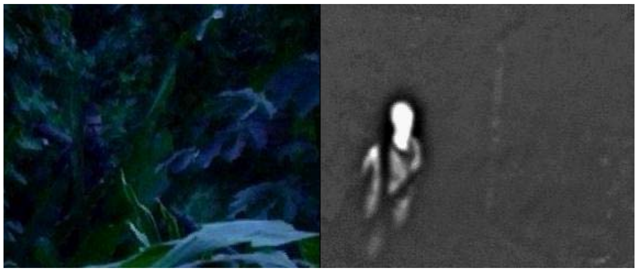

Figure 2: SIM-LL Viz and IR SBS display

The SIM multi-sensor video data together with the observer scanpath data from our experiments is publicly available at Scanpaths.org [23]. The COM sequences (Figures 3 and 4) showed a group of people in a clearing of trees, interacting, some behaving suspiciously (carrying and leaving a rucksack, using mobile phones, running etc.) The sequences were shown separately; with the Viz display on the left-hand side and IR on the right-hand side, next to each other, and each display (Viz and IR) being 288 x 240 pixels with a 1 pixel black line between them.

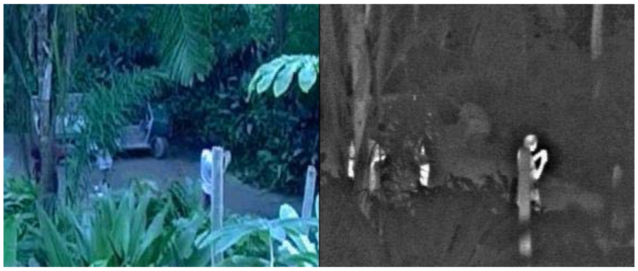

Figure 3: COM-HL Viz and IR SBS display

The sequences were split into sections on the basis of their content. In the SIM sequences, there were three sections, each being one period that the soldier was visible (Section 1 walking down the path; Section 2 walking across the clearing left-to-right; Section 3 walking right-to-left). In the COM-HL sequence there were two sections: these corresponded with the two times the target individual was in clearing, whilst in COM-LL there were four, again when the target was in the clearing.

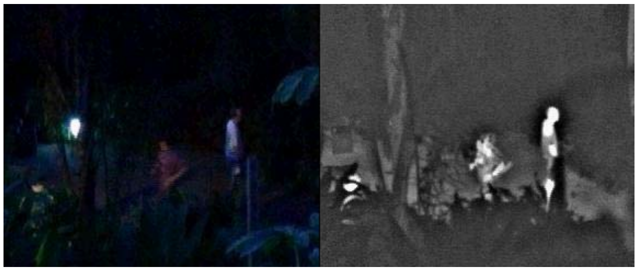

Figure 4: COM-LL Viz and IR SBS display

2.4 Procedure

Participants were asked to attend three sessions, each session consisting of the same experimental conditions. In the SIM scenarios they were asked to track the soldier throughout the sequence (even whilst obscured), as well as to press 'space' when the soldier passed certain features of the scene. In the COM scenarios they were asked to track the individual who was positioned approximately centrally to the screen and wearing a white t-shirt. They were also asked to press 'B' when they saw anyone carrying a rucksack, and 'N' when they saw anyone using a camera or mobile phone (COM-HL), or running or disguising themselves (COM-LL).

2.5 Data Analysis

In order to analyse how accurately participants tracked the given target from the scanpath data obtained, gaze locations (the eye fixation data) had to be compared with pre-drawn ground truth 'target boxes'. These were rectangular target boxes drawn around the soldier (target to be tracked), that were created manually using a Matlab toolbox we have designed, that can be used to delineate rectangles throughout a sequence. Targets were drawn at least every 15 frames where possible; when the tracking target was not visible for longer periods on the screen estimations were made. Once the target rectangles were drawn, the in-between rectangles were calculated by interpolation and then visually inspected. Where necessary, these were individually readjusted.

Raw eye fixation data were taken and compiled so that they could be compared with the target boxes previously created. This is shown in Figure 5. Once it was known in which frame each recorded gaze point was located, a direct evaluation could be made with the target overlays. For each display modality in each task an 'accuracy ratio' was calculated by dividing the number of gaze points located inside the target map by the total number of gaze fixations recorded.

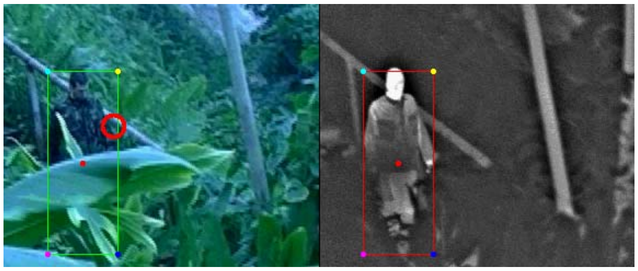

Figure 5: Viz and IR SBS display with target and fixation

The ClearView program also supplied data on how valid each raw fixation was for each eye. This ranged from '0' (definitely certain that a particular fixation belonged to a particular eye) to '3' (very uncertain that a gaze point corresponds to an eye), with '4' meaning that no eye was detected. In the current study, only fixation data with a validity of '0' for both eyes was used. The eye fixation points of the two eyes were then averaged for every recorded pair of gaze fixations. This provided the most valid data, averaged to accommodate any variance caused by 'drifting' artefacts, which are usually inversely symmetrical.

3 Results

The current results are presented in four sections for each scenario used. They each comprise one analysis of the overall percentage of the time participants viewed the Viz or IR display within the SBS condition, one analysis of tracking accuracy within the SBS display, and one comparison of the display results with the other fused and input displays. For brevity, the final comparison with the other fusion and input methods does not have a figure included.

3.1 Simple - High Luminance (SIM-HL)

The SIM-HL overall looking data are shown in Figure 6. A three-way dependent measures ANOVA revealed a close-to-significant main effect of display (F(1, 9) = 4.17, p = .071), but not session, nor section and no interactions. This suggests that IR was close to being viewed overall more than Viz.

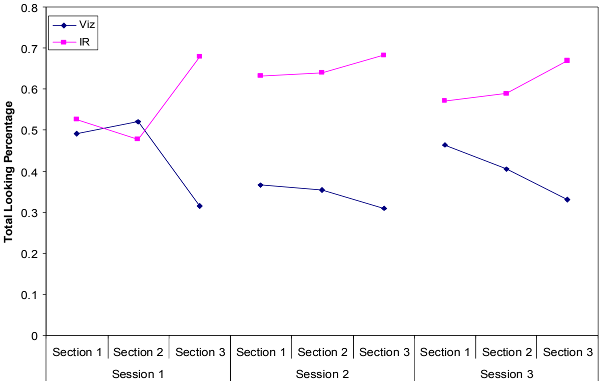

Figure 6: ANOVA results of overall SIM-HL data

The SBS accuracy results, as shown in Figure 7, were analysed using an ANOVA, which revealed main effects of display (F(1, 9) = 10.6, p = .010), session (F(2, 18) = 8.53, p = .002), and section (F(2, 18) = 9.73, p = .001), although no interactions were found. Bonferroni analyses of the latter results indicated that Session 2 was significantly greater than Session 1 (p = .008), whilst Section 3 was significantly lower than Sections 1 (p = .009) and 2 (p = .020). These results indicate that when the SBS data were split into sections, the IR display led to significantly greater accuracy than Viz, Session 2 had greater accuracy than the other Sessions, whilst the third section had lower accuracy than Sections 1 and 2, as was found in the previous analysis.

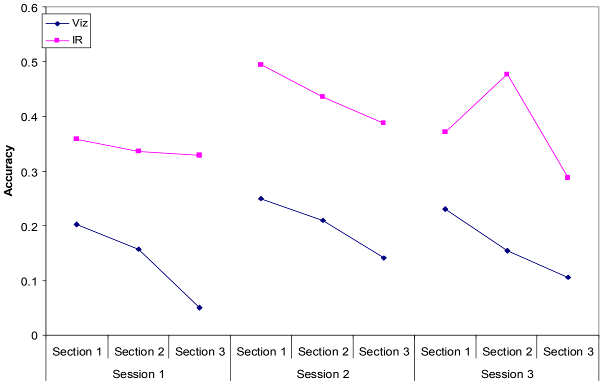

Figure 7: ANOVA results of SIM-HL accuracy data

Pairwise comparisons of the SBS accuracy data with the other inputs and fusion methods (see also [1]) reveal the Viz+IR display to be significantly lower than the AVE fused display (p = .010), with AVE having greatest accuracy, but no other differences were found.

3.2 Simple – Low Luminance (SIM-LL)

The overall SIM-LL SBS data (Figure 8) was tested again with an ANOVA, which revealed a main effect of display (F(1, 9) = 122, p < .001), showing IR to be significantly favoured over Viz, although no effect of session, nor section, although there were significant interactions between display and session (F(2, 18) = 5.83, p = .011) and between display and section (F(2, 18) = 16.0, p < .001).

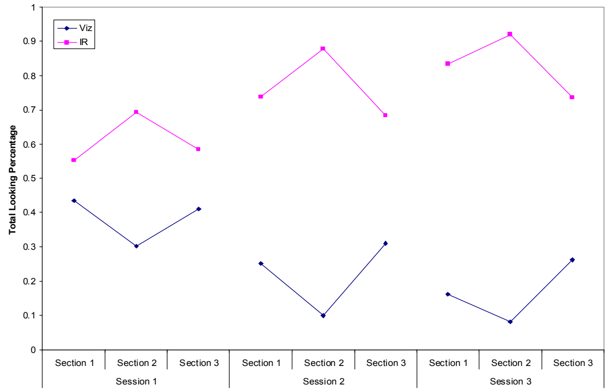

Figure 8: ANOVA results of overall SIM-LL data

Tukey post hoc testing of the display by session interaction revealed no significant difference between Viz and IR in Session 1, but IR was significantly greater than Viz in Sessions 2 (HSD = .449, p = .05) and 3 (HSD = .560, p = .01), showing that participants were learning across sessions to use the IR more and Viz less. The ‘display by section’ interaction revealed significant differences between Viz and IR in Sessions 1 (HSD = .296, p = .01), 2 (HSD = .378, p = .001) and 3 (p = .01), with the stronger effect in Section 2. This indicates that participants were favouring IR more in the second section than Sections 1 and 3, although IR was always favoured significantly more than Viz.

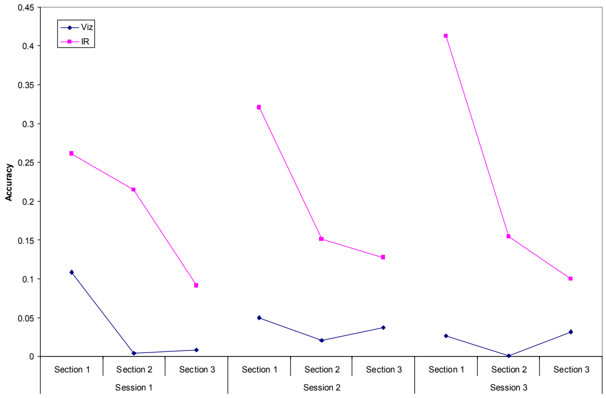

Figure 9: ANOVA results of SIM-LL accuracy data

The SBS accuracy results in Figure 9 showed IR to be more accurate than Viz, which was supported by ANOVA analysis which revealed a main effect of display (F(1, 9) = 252, p < .001), no effect of session, but there was a main effect of section (F(2, 18) = 47.4, p < .001), as well as an interaction between display and section (F(2, 18) = 16.4, p < .001). Bonferroni testing of the section effect revealed significant differences between Sections 1 and 2 (p < .001) and Sections 1 and 3 (p < .001).

Tukey testing of the two-way interaction revealed IR to be significantly greater than Viz in Section 1 (HSD = .158, p = .01) and Section 2 (p = .01), but not Section 3. In addition, IR dropped significantly between Sections 1 and 2 (HSD = .127, p = .05) whilst Viz did not significantly change between these sections.

Statistical comparison of the SIM-LL data with the other displays (see also [1]) showed no significant differences between the Viz+IR SBS display and the other fused and input displays, with DWT fusion having greatest accuracy within the dataset.

3.3 Complex – High Luminance (COM-HL)

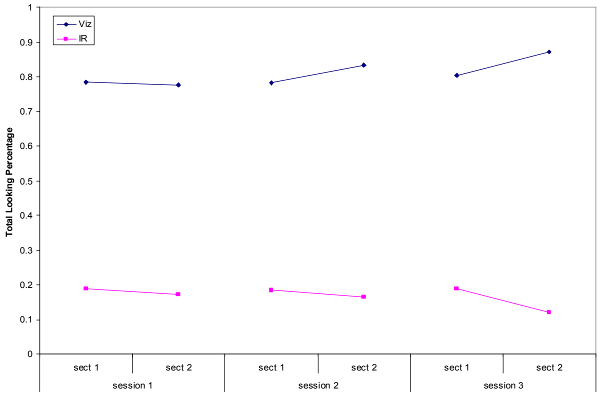

Figure 10: ANOVA results of overall COM-HL data

An ANOVA analysis of the COM-HL data, plotted in Figure 10, revealed a main effect of display (F(1, 9) = 29.9, p < .001), but neither session nor section and no interactions. This indicates that the Viz display was preferred over the IR.

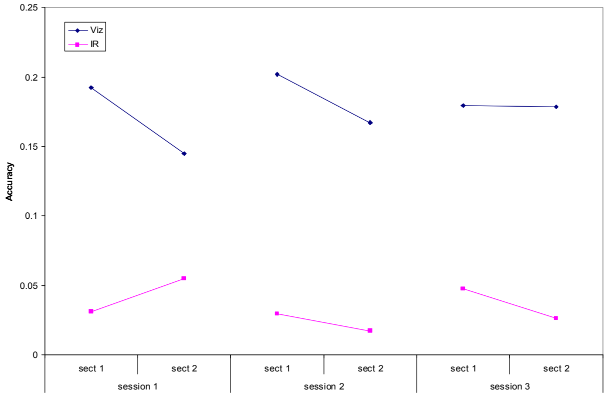

Figure 11: ANOVA results of COM-HL accuracy data

The accuracy scores in Figure 11 reveal a similar pattern, as in Figure 10, with an ANOVA revealing Viz to be significantly more accurate than IR (F(1, 9) = 14.1, p =

.005), but no main effects of session or section and no interactions.

Comparison of this data with the other displays (see also [2]) revealed the Viz+IR inputs to be significantly lower than Viz (p < .001), IR (p = .020), AVE (p = .012) and DT-CWT (p = .001), suggesting that there was much detriment to viewing accuracy in the COM-HL sequence when using the SBS display. In this scenario, Viz had greatest accuracy.

3.4 Complex – Low Luminance (COM-LL)

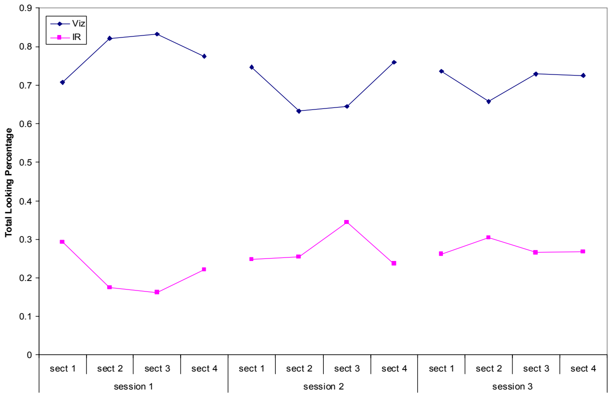

Figure 12: ANOVA results of overall COM-LL data

An ANOVA analysis of the COM-LL data, as shown in Figure 12, revealed a main effect of display (F(1, 9) = 14.5, p = .004), indicating that again participants favoured the Viz over IR display, whilst there were no effects of session or section, and no interactions.

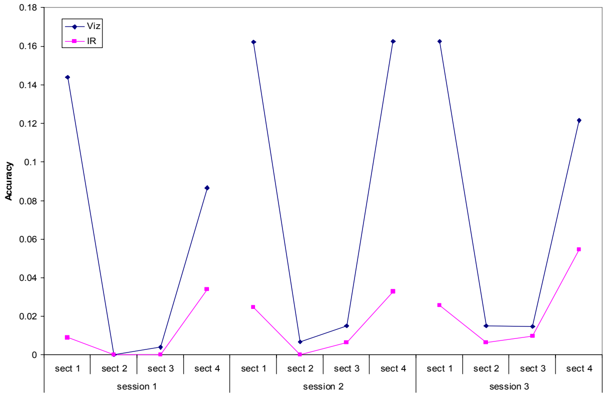

Figure 13: ANOVA results of COM-LL accuracy data

In contrast, the accuracy data for the COM-LL dataset (Figure 13) revealed main effects of display (F(1, 9 = 5.90, p = .38) showing Viz to lead to significantly higher accuracy than IR, and main effects of session (F(2, 18) = 4.36, p = .028), and session (F(3, 27) = 13.8, p = .001), as well as an interaction between display and section (F(3, 27) = 5.94, p = .025). Bonferroni testing of the main effect of session indicated that Session 2 was significantly greater in accuracy than Session 1 (p = .046), whilst the main effect of section revealed Sections 2 and 3 to be significantly lower than Sections 1 (p = .018) and 4 (p = .028; p = .019). The interaction was revealed through Tukey HSD testing to lie between Sections 1 and 2, with Viz significantly falling between these sections but IR not (HSD = .147, p = .05).

The comparison between Viz+IR and the other displays (see also [2]) showed the SBS display to be significantly lower than the Viz display (p = .018), with no other differences found, whilst DT-CWT had led to greatest accuracy.

3.5 Switching Behaviour

The pattern of switching from Viz to IR and IR to Viz was also calculated for each participant, ignoring the separate sections within each scenario. These were then analysed with a three-way repeated measures ANOVA comparing switch direction (Viz-IR; IR-Viz), the four scenarios and the three sessions. This revealed a main effect of switch direction (F(1, 9) = 6.16, p = .035), indicating that participants switched from IR to Viz more than Viz to IR, a main effect of scenario (F(3, 27) = 14.8, p < .001), with COM-HL lower than the rest, and a main effect of session (F(2, 18) = 29.6, p < .001), with Session 1 having significantly more switching than Sessions 2 and 3. It should be noted that the scenario results was somewhat confounded by the fact that the whole sequence switching data could not be used, only the first 1305 frames.

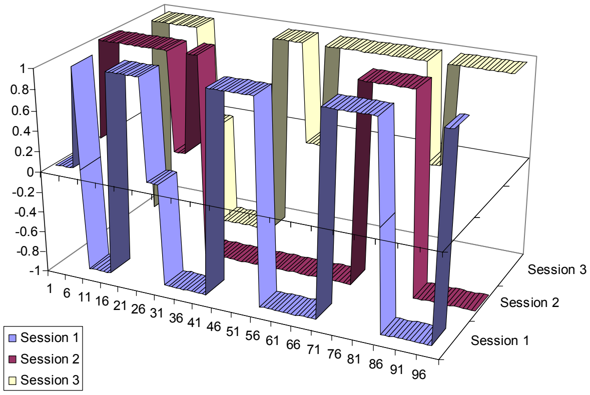

Figure 14: Participant 1 switching behaviour

Figure 14 shows the switching behaviour for Participant 1, in all three sessions of the SIM-HL scenario, across the first 100 frames of the sequence. A score of 1 indicates viewing the Viz (left-hand) display, and -1 indicates viewing IR (right-hand) display, whilst a score of 0 shows that the participant looked at neither of the Viz and IR sections of the SBS display but at the surrounding black area of the screen.

4 Discussion

The current work has produced a range of interesting and varied findings. The first issue of interest to note is the difference in favouring either the Viz or IR in SBS displays in the simple and complex tasks. It would seem that when the task was simple, little contextual information was required in order to accurately track the figure person? (as well as performing the secondary task).

However, when there were several people interacting, and the secondary tasks were more cognitively demanding, participants relied more on the Viz display. One possible issue affecting this finding is that the task demands of the complex scenarios may have necessitated participants use the Viz display, as the colour of the person's T-shirt would not be available in the IR. Nevertheless, it is surprising that participants still favoured Viz even in the COM-LL sequence, as there was little tracking information to be obtained from that display. This might indicate the participants were more focused on the secondary task in this scenario; further analysis of the datasets compared with secondary task performance might establish this.

A second related point is that the participants in the COM-LL scenario improved in accuracy over the three sessions. It might therefore be that whilst they relied upon the Viz display more than the IR (as overall choice of display did not vary across sessions), they were learning, perhaps from viewing the fused displays, where to look to track the figure.

Thirdly, the results indicate that in the SIM-LL data participants were learning across repeated viewings to use the IR display more and the Viz display less. This suggests that, unlike our previous findings [1, 2], participants can learn across sessions when a SBS display is used. In addition, the advantage of using the IR display in terms of accuracy was shown to drop in the final section of the SIM-LL sequence, suggesting that, despite the fact that participants were still favouring the IR display overall in Section 3, they were hindered in tracking the target in this section.

The COM-HL results are of note due to the fact that it is only in this scenario that the Viz+IR accuracy scores are significantly below the majority of the other displays (fused or input). The Viz+IR display was generated by halving the size of each of the two inputs, and placing them side by side. Therefore, the inputs on their own, as well as the fused datasets, were twice the size of either one of the Viz+IR. In turn, each of the target boxes in the SBS was a quarter of the area of the original, so even with two target boxes included (and remembering that participants can only look at either the Viz or IR target at any one time) there was still half the 'target space' in the SBS condition. It is therefore surprising that there is only one scenario wherein the Viz+IR scores much worse than most of the other displays. This perhaps suggests that a Viz+IR SBS display may still serve as a fairly reliable display dependent on task.

In terms of total accuracy, the Viz part of each SVS display can be compared with the Viz input accuracy, and IR part of the SBS with the IR input. For the SIM-HL sequence, Viz input scored .593, whilst Viz-SBS scored .167. Likewise, IR input scored .645 and IR-SBS scored .382. In SIM-LL, Viz-input was .190 whilst Viz-SBS was .032 and IR-input was .236 with IR-SBS .204. COM-LL Viz-input accuracy came to .363; Viz-SBS equalled .177, with IR-input .312 and IR-SBS .034. Finally, COM-LL Viz-input came to .152, Viz-SBS came to .076, IR-input was .167 and IR-SBS came to .017. These data, although not statistically validated, suggest that the individual SBS displays are much lower than the inputs, as could be concluded from the fact that they are half the size. One notable exception is the SIM-LL IR-SBS, which led to comparable accuracy to the IR-input. It must also be noted that any SBS accuracy score has to be 'shared' between both the Viz and IR sides of the display, which would mean that participants would have to almost exclusively choose one side in order to obtain anything like as close an accuracy score as one of the input displays. This is indeed the case in the SIM-IR results, as detailed in Section 3.2.

Finally, the switching analysis suggests that participants were learning to swap viewing sides less across repeated viewings. In addition, they were making more switches from IR to Viz than vice versa, which is surprising given that the SIM tasks favoured the IR and COM favoured Viz. What it does perhaps suggest is that participants would usually make an initial IR fixation in a given period of viewing, and then either stay there for longer or switch to the Viz, dependent on which provided best usage for the given task.

5 Conclusion

The current work has found several interesting and valuable results relating to the use of SBS displays. Such displays have been shown to perform poorly in applied settings [6], whilst the process of switching attention from one source of information to another can lead to increased reaction times and confused responses [5]. The current results support this to some extent by finding participant accuracy to be somewhat reduced for each of the individual displays compared with showing the inputs alone. However, when the accuracy of the two sides of the SBS display was combined, then accuracy only fell below most inputs and other fusion methods in the COM-HL task. As the very nature of the SBS display leads participants to switch between sides, it is not surprising that the complex task led to poorer accuracy for the SBS display, as loss of attention would have a larger detriment in this scenario.

One prospective line of further investigation involves the use of the current dataset to model human scanpath behaviour when presented with a SBS or fused display. Such an approach could be used to create standardised viewing metrics for various fusion systems that could predict what aspects of a fused, SBS or input display would attract visual attention. This in turn could potentially allow for online adjustment of the display being used in order to maximise usability.

In general, it can be posited that use of SBS displays will not necessarily lead to worse performance than either input alone, but dependent upon the task, performance can drop. In addition, as there is no detriment found in the current experiments from using a fused over SBS display, and given this, there can be advantages to using a fused video. Moreover, choosing to combine information into a single output should also be considered when deciding upon the kind of display to adopt, dependent upon the task being undertaken.

Acknowledgements

This project is funded by the UK MOD Data and Information Fusion Defence Technology Centre (DIF-DTC). We would like to thank the Eden Project in Cornwall, UK, for their support in this project.

References

[1] T.D. Dixon et al., *Scanpath analysis of fused multi-sensor images with luminance change: A pilot study*, 9th International Conference on Information Fusion (Fusion 2006), Image Fusion Assessment, Florence, IT, 10-13 July 2006.

[2] T.D. Dixon et al., *Assesment of fused videos using scanpaths: A comparison of data analysis methods,* Spat. Vis., Vol 20, April 2007 (to appear).

[3] D. J. Hawkes, et al., Preliminary work on the integration of SPECT images with the aid of registered MRI images and an MR derived neuro-anatomical atlas. In K. H. Hoehne, H. Fuchs, and S. M. Pizer, editors, 3D Imaging in Medicine: Algorithms, Systems, Applications, pages 241–252. Springer Verlag, 1990.

[4] R Stokking. Integrated Visualization of Functional and Anatomical Brain images. PhD thesis, University of Utrecht, Utrecht, 1998.

[5] J. Rabin and R. Wiley, *Switching from forward-looking infrared to night vision goggles: Transitory effects on visual resolution,* ASEM, Vol. 65, No. 4, pp 327-329 April 1994.

[6] H. Amthauer et al. *Value of image fusion using single photon emission computed tomography with integrated low does computed tomography in comparison with retrospective voxel-based method in neuroendocrine tumours,* Eur. Radiol. Vol. 15, pp 1456-1462, 2005.

[7] A. Loza, et al. *Methods of fused image analysis and assessment.* In Proceedings of the Advanced Study Institute Conference (NATO-ASI 2005): Multisensor Data and Information Processing for Rapid and Robust Situation and Threat Assessment, Bulgaria, May, 2005.

[8] G. Simone, et al. *Image fusion techniques for remote sensing applications.* Information Fusion, Vol. 3, pp3-15, 2002.

[9] N. Kingsbury, *Image processing with complex wavelets.* In Wavelets: The Key to Intermittent Information, B. Silverman and J. Vassilicos, Eds. Oxford University Press, USA, pp165–185, 1999.

[10] S.G. Nikolov et al. *Wavelets for image fusion*. In, *Wavelets in Signal and Image Analysis*, Computational Imaging and Vision Series, A. Petrosian and F. Meyer, eds. pp213–244 Kluwer Academic Publishers, Dordrecht, Netherlands, 2001.

[11] T Dixon et al. *Quality assessment of false colored fused displays.* JSID, Vol. 14, No. 10, pp883-894, October 2006.

[12] T. Dixon et al. *Methods for the assessment of fused images,* ACM TAP, Vol. 3, No. 3, pp309-332, 2006.

[13] A. Toet, and E.M. Franken. *Perceptual evaluation of different image fusion schemes*. Disp. Vol. 24, No. 1 pp25-37, February, 2003.

[14] A. Toet., et al. *Fusion of visible and thermal imagery improves situational awareness*. Disp., Vol. 18 pp85-95, 1997.

[15] R.H.S. Carpenter. *Movements of the Eye.* Pion, London, 1977.

[16] J. M. Henderson and A. Hollingworth. *High-level scene perception.* ARP, Vol. 50 pp243-271, 1999.

[17] G. Underwood. *Eye fixations on pictures of natural scenes: Getting the gist and identifying the components.* In Cognitive Processes in Eye Guidance, G. Underwood Ed, Oxford University Press, New York, 2005.

[18] S.B. Hutton and D. Tegally. *The effects of dividing attention on smooth pursuit eye tracking.* EBR, Vol. 163, pp306-313, 2005.

[19] W.K. Krebs et al., *Comparing behavioral receiver operating characteristic curves to multidimensional matched filters,* Opt. Eng., Vol. 40, No. 9, pp1818-1826, 2001.

[20] Y. Lanir, *Comparing multispectral image fusion methods for a target detection task,* Masters Thesis, Ben-Gurion University of Negev, June 2005.

[21] Tobii Technology: http://www.tobii.se/

[22] J. J. Lewis, et al. *The Eden Project multi-sensor data set.* Technical report TR-UoB-WS-Eden-Project-Data-Set, available from The Online Resource for Research in Image Fusion (ImageFusion.org), 2006.

[23] The Online Archive of Scanpath Data (Scanpaths.org), 2006.